DOCUMENT RESUME

ED 328 611 TM 016 142

AUTHOR Wiggins, Grant
TITLE The Case for Authentic Assessment. ERIC Digest.
INSTITUTION American Institutes for Research, Washington, DC.; ERIC Clearinghouse on Tests, Measurement, and Evaluation, Washington, DC.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO EDO-TM-90-10
PUB DATE Dec 90
CONTRACT RI88062003
NOTE 4p.
PUB TYPE Information Analyses - ERIC Information Analysis Products (071)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Testing; Cost Effectiveness; *Educational Assessment; Elementary Secondary Education; Nontraditional Education; Public Opinion; Standardized Tests; *Test Use; Test Validity
IDENTIFIERS *Authentic Assessment; California Assessment Program; *Direct Assessment; ERIC Digests; Indirect Assessment

ABSTRACT

Based on material prepared for the California Assessment Program, an argument in favor of authentic assessment is presented, and authentic assessment is contrasted with traditional standardized tests. An assessment is authentic when student performance on intellectual tasks is directly examined. Comparatively, traditional assessment relies on indirect or proxy items. Issues addressed include cost, effort and time requirements, and public suspicions regarding the objectivity of authentic assessment. It is contended that a move toward more authentic tasks and outcomes improves teaching and learning. In authentic assessments, students have greater clarity about their obligations and are asked to master more engaging tasks, and teachers can see assessment results as meaningful and useful for improving instruction. Conventional testing is probably adequate if the aim is to monitor performance. However, tests must be composed of exemplary tasks, criteria, and standards if the goal is to improve performance across the board. A five-item list of additional reading materials is included. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED328611

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

THE CASE FOR AUTHENTIC ASSESSMENT

by Grant Wiggins, CLASS

TM016142

The Case for Authentic Assessment

by Grant Wiggins, CLASS

Mr. Wiggins, a researcher and consultant on school reform issues, is a widely-known advocate of authentic assessment in education. This digest is based on materials that he prepared for the California Assessment Program.

What is Authentic Assessment?

Assessment is authentic when we directly examine student performance on worthy intellectual tasks. Traditional assessment, by contract, relies on indirect or proxy 'items'–efficient, simplistic substitutes from which we think valid inferences can be made about the student's performance at those valued challenges.

Do we want to evaluate student problem-posing and problem-solving in mathematics? Experimental research in science? Speaking, listening, and facilitating a discussion? doing document-based historical inquiry? thoroughly revising a piece of imaginative writing until it 'works' for the reader? Then let our assessment be built out of such exemplary intellectual challenges.

Further comparisons with traditional standardized tests will help to clarify what 'authenticity' means when considering assessment design and, use:

- Authentic assessments require students to be effective performers with acquired knowledge. Traditional tests tend to reveal only whether the student can recognize, recall or 'plug in' what was learned out of context. This may be as problematic as inferring driving or teaching ability from written tests alone. (Note, therefore, that the debate is not "either-or": there may well be virtue in an array of local and state assessment instruments as befits the purpose of the measurement).
- Authentic assessments present the student with the full array of tasks that mirror the priorities and challenges found in the best instructional activities: conducting research; writing, revising and discussing papers; providing an engaging oral analysis of a recent political event; collaborating with others on a debate, etc. Conventional tests are usually limited to paper-and-pencil, one-answer questions.
- Authentic assessments attend to whether the student can craft polished, thorough and justifiable answers, performances or products. Conventional tests typically only ask the student to select or write correct responses–irrespective of reasons. (There is rarely an adequate opportunity to plan, revise and substantiate responses on typical tests, even when there are open-ended questions). As a result,
- Authentic assessment achieves validity and reliability by emphasizing and standardizing the appropriate *criteria* for scoring such (varied) products; traditional testing standardizes objective 'items' and, hence, the (one) *right answer* for each.
- "Test validity" should depend in part upon whether the test simulates real-world 'tests' of ability. Validity on most multiple-choice tests is determined merely by matching items to the curriculum content (or through sophisticated correlations with other test results).
- Authentic tasks involve 'ill-structured' challenges and roles that help students rehearse for the complex ambiguities of the 'game' of adult and professional life. Traditional tests are more like drills, assessing static and too-often arbitrarily discrete or simplistic elements of those activities.

Beyond these technical considerations the move to reform assessment is based upon the premise that assessment should primarily support the needs of learners. Thus, secretive tests composed of proxy items; and scores that have no obvious meaning or usefulness undermine teachers' ability to improve instruction and students' ability to improve their performance. We rehearse for and teach to authentic tests–think of music and military training–without compromising validity.

The best tests always teach students and teachers alike the kind of work that most matters; they are enabling and forward-looking, not just reflective of prior teaching. In many colleges and all professional settings the essential challenges are known in advance–the upcoming report, recital, Board presentation, legal case, book to write, etc. Traditional tests, by requiring complete secrecy for their validity, make it difficult for teachers and students to rehearse and gain the confidence that come from knowing their performance obligations. (A known challenge also makes it possible to hold all students to higher standards).

Why do we need to invest in these labor-intensive forms of assessment?

While multiple-choice tests can be valid indicators or predictors of academic performance, too often our tests mislead students and teachers about the kinds of work that should be mastered. Norms are not standards; items are not real problems; right answers are not rationales.

What most defenders of traditional tests fail to see is that it is the *form*, not the content of the test that is harmful to learning; demonstrations of the technical validity of standardized tests should not be the issue in the assessment reform debate. Students come to believe that learning is cramming; teachers

TM016142

come to believe that tests are after-the-fact, imposed nuisances composed of contrived questions–irrelevant to their intent and success. Both parties are led to believe that right answers matter more than habits of mind and the justification of one's approach and results.

A move toward more authentic tasks and outcomes thus improves teaching and learning: students have greater clarity about their obligations (and are asked to master more engaging tasks), and teachers can come to believe that assessment results are both meaningful and useful for improving instruction.

If our aim is merely to monitor performance than conventional testing is probably adequate. If our aim is to improve performance across the board than the tests must be composed of exemplary tasks, criteria and standards.

Won't authentic assessment be too expensive and time-consuming?

The costs are deceptive: while the scoring of judgment-based tasks seems expensive when compared to multiple-choice tests (about $2 per student vs. 1 cent) the gains to teacher professional development, local assessing, and student learning are many. As states like California and New York have found (with their writing and hands-on science tests) significant improvements occur locally in the teaching and assessing of writing and science when teachers become involved and invested in the scoring process.

If costs prove prohibitive, sampling may well be the appropriate response–the strategy employed in California, Vermont and Connecticut in their new performance and portfolio assessment projects. Whether through a sampling of many writing genres, where each student gets one prompt only; or through sampling a small number of all student papers and school-wide portfolios; or through assessing only a small sample of students, valuable information is gained at a minimum cost.

And what have we gained by failing to adequately assess all the capacities and outcomes we profess to value simply because it is time-consuming, expensive, or labor-intensive? Most other countries routinely ask students to respond orally and in writing on their major tests–the same countries that outperform us on international comparisons. Money, time and training are routinely set aside to insure that assessment is of high quality. They also correctly assume that high standards depend on the quality of day-to-day local assessment–further offsetting the apparent high cost of training teachers to score student work in regional or national assessments.

Will the public have any faith in the objectivity and reliability of judgment-based scores?

We forget that numerous state and national testing programs with a high degree of credibility and integrity have for many years operated using human judges:

- the New York Regents exams, parts of which have included essay questions since their inception–and which are scored locally (while audited by the state);
- the Advanced Placement program which uses open-ended questions and tasks, including not only essays on most tests but the performance-based tests in the Art Portfolio and Foreign Language exams;
- state-wide writing assessments in two dozen states where model papers, training of readers, papers read 'blind' and procedures to prevent bias and drift gain adequate reliability;
- the National Assessment of Educational Progress (NAEP). the Congressionally-mandated assessment, uses numerous open-ended test questions and writing prompts (and successfully piloted a hands-on test of science performance)
- newly-mandated performance-based and portfolio-based state-wide testing in Arizona, California, Connecticut, Kentucky, Maryland, and New York.

Though the scoring of standardized tests is not subject to significant error, the procedure by which items are *chosen*, and the manner in which norms or cut-scores are established is often quite subjective–and typically immune from public scrutiny and oversight.

Genuine accountability does not avoid human judgment. We monitor and improve judgment through training sessions, model performances used as exemplars, audit and oversight policies as well as through such basic procedures as having disinterested judges review student work 'blind' to the name or experience of the student–as occurs routinely throughout the professional, athletic and artistic worlds in the judging of performance.

Authentic assessment also has the advantage of providing parents and community members with directly observable products and understandable evidence concerning their students' performance; the quality of student work is more discernible to laypersons than when we must rely on translations of talk about stanines and renorming.

Ultimately, as the researcher Lauren Resnick has put it, What you assess is what you get; if you don't test it you won't get it. To improve student performance we must recognize that essential intellectual abilities are falling through the cracks of conventional testing.

Additional Reading

Archbald, D. & Newmann, F. (1989) "The Functions of Assessment and the Nature of Authentic Academic Achievement," in Berlak (ed.) *Assessing Achievement: Toward the Development of a New Science of Educational Testing*. Buffalo, NY: SUNY Press.

Frederiksen, J. & Collins, A. (1989) "A Systems Approach to Educational Testing," *Educational Researcher*, 18, 9 (December).

National Commission on Testing and Public Policy (1990) *From Gatekeeper to Gateway: Transforming Testing in America*. Chestnut Hill, MA: NCTPP, Boston College.

Wiggins, G. (1989) "A True Test: Toward More Authentic and Equitable Assessment," *Phi Delta Kappan*, 70, 9 (May).

Wolf, D. (1989) "Portfolio Assessment: Sampling Student Work," *Educational Leadership* 46, 7, pp. 35-39 (April.)

This publication was prepared with funding from the Office of Educational Research and Improvement, U.S. Department of Education under contract number R-88-062003. The opinions expressed in this report do not necessarily reflect the position or policies of OERI or the Department of Education. ***Permission is granted to copy and distribute this ERIC/TM Digest.***